Zhongyang MAO, Jiahuan GENG, Faping LU, Wenbiao TIAN, Jiafang KANG, Yaozong PAN, 2026. An adaptive meta-reinforcement learning scheme for maritime dynamic spectrum access and service scheduling. *ENGINEERING Information Technology & Electronic Engineering*, 27(8):260047. https://doi.org/10.1631/ENG.ITEE.2026.0047

ENGINEERING Inform Technol Electron Eng

An adaptive meta-reinforcement learning scheme for maritime dynamic spectrum access and service scheduling

Key words: Dynamic spectrum access; Adaptive meta-reinforcement learning; Service scheduling; Maritime communication

Corresponding author: Jiahuan GENG
E-mail: echo_HAU@163.com
ORCID: https://orcid.org/0009-0004-9031-0690

Motivation

- Maritime communications face scarce spectrum, highly dynamic channel conditions, and heterogeneous service requirements with urgent and normal traffic, yet conventional deep reinforcement learning (DRL) policies trained for specific environments generalize poorly to unknown scenarios and lack fast, reliable adaptation, necessitating an intelligent framework that can swiftly learn effective spectrum access and service scheduling across diverse, non-stationary maritime scenarios.

ENGINEERING Inform Technol Electron Eng

Main idea

- A maritime system model is constructed by incorporating realistic channel occupancy patterns and a dual-queue mechanism for urgent and normal packets.
- A meta-reinforcement learning-based spectrum access and service scheduling algorithm is proposed to overcome the poor generalization of traditional DRL.
- An adaptive meta-reinforcement learning spectrum access and service scheduling algorithm is introduced to accelerate convergence and achieve efficient adaptation to complex, varying maritime scenarios.

ENGINEERING Journal Technical Electron Eng

Method

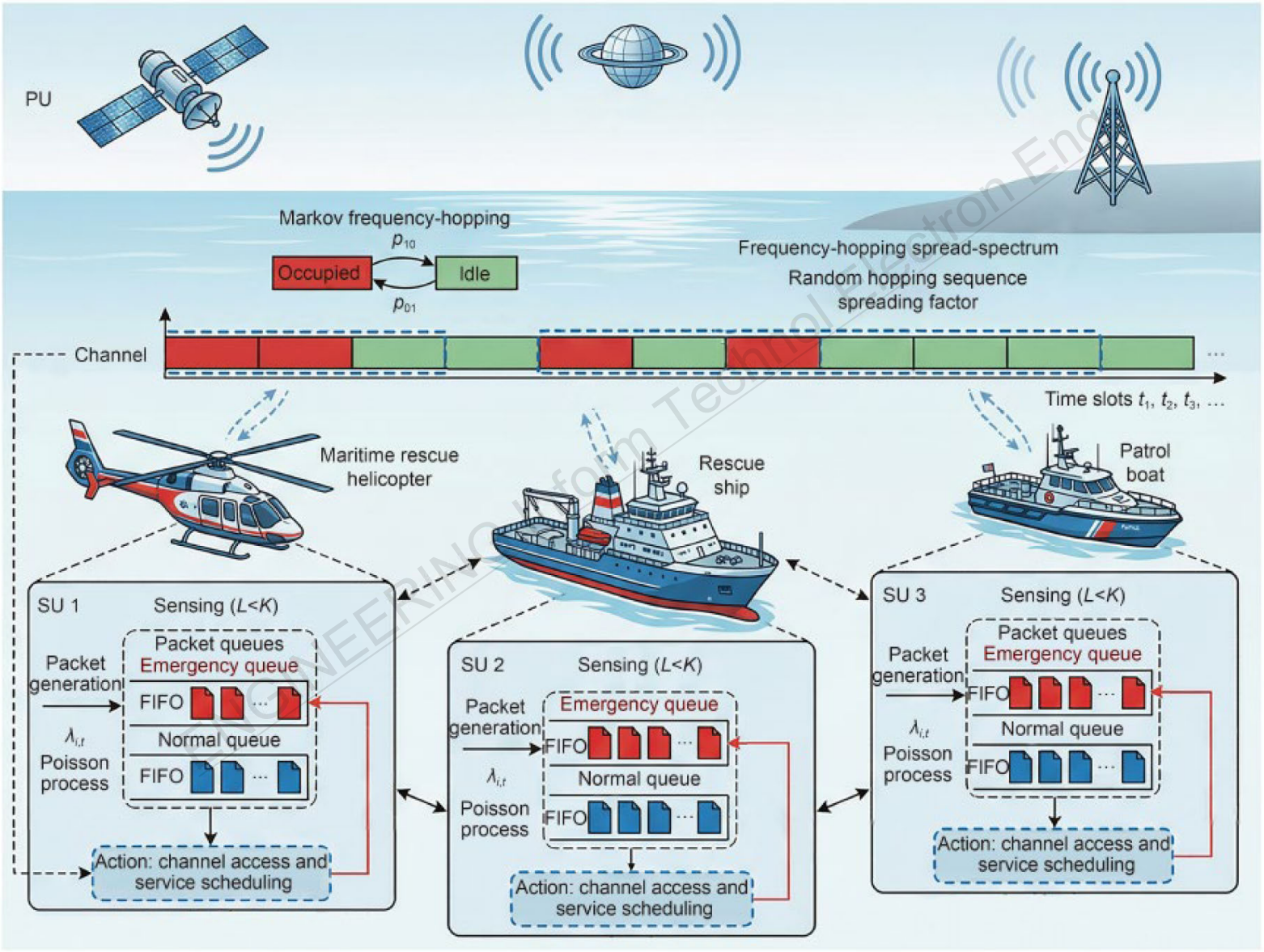


Fig. 1 Architecture of dynamic spectrum access and service scheduling for maritime emergency rescue

Method

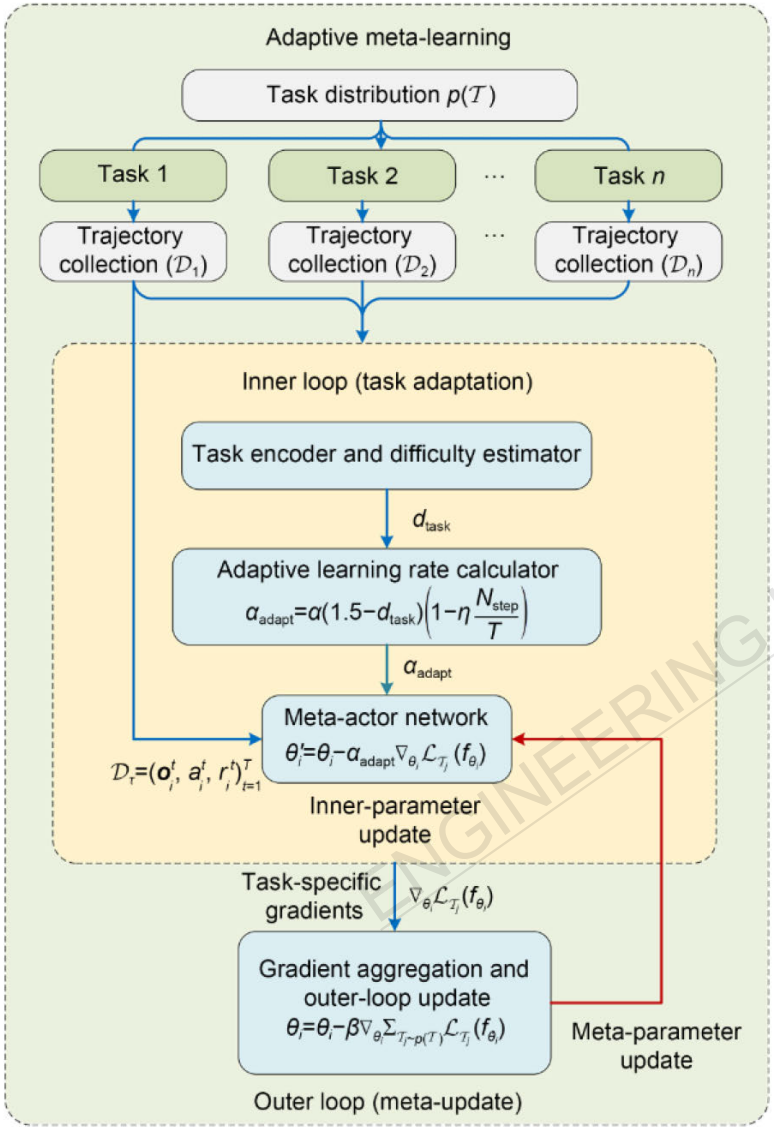


Fig. 2 Framework of the adaptive meta-learning algorithm model

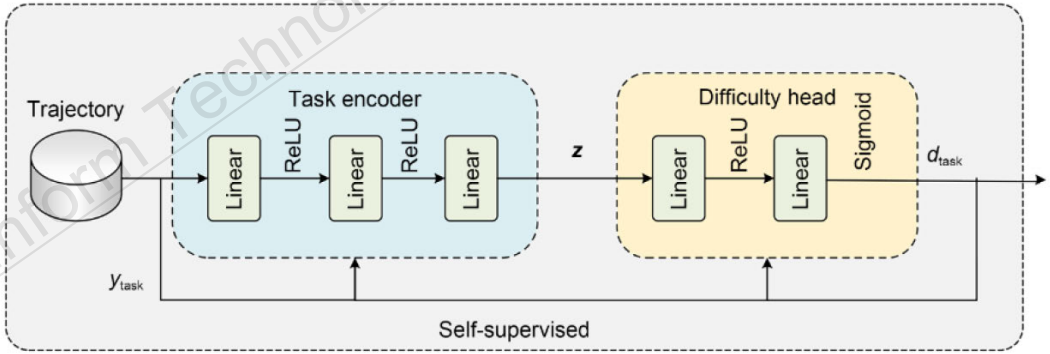


Fig. 3 Framework of the task perception and difficulty estimation module

ENGINEERING Inform Technol Electron Eng

Results

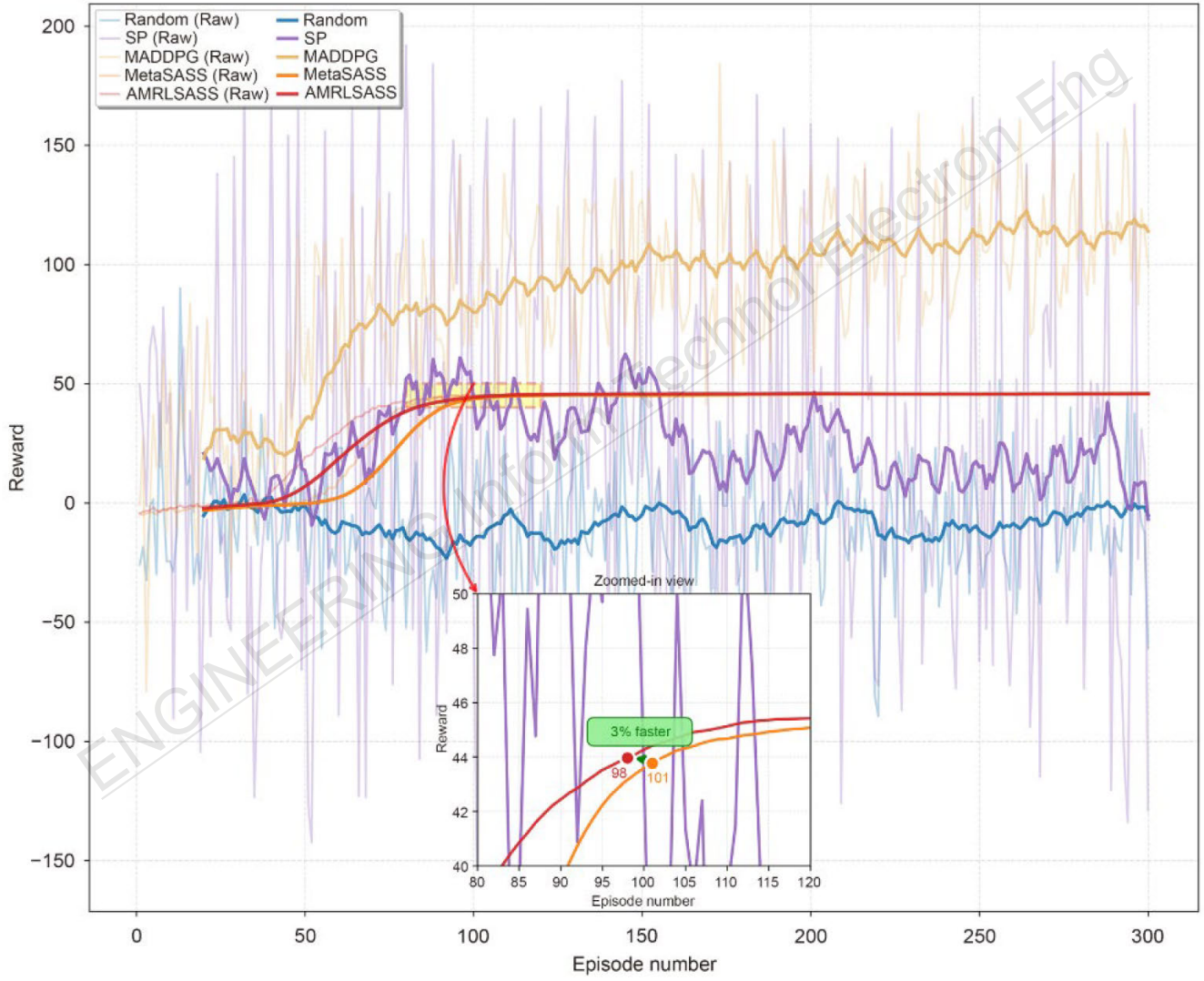


Fig. 4 Training curves of the five algorithms

Results

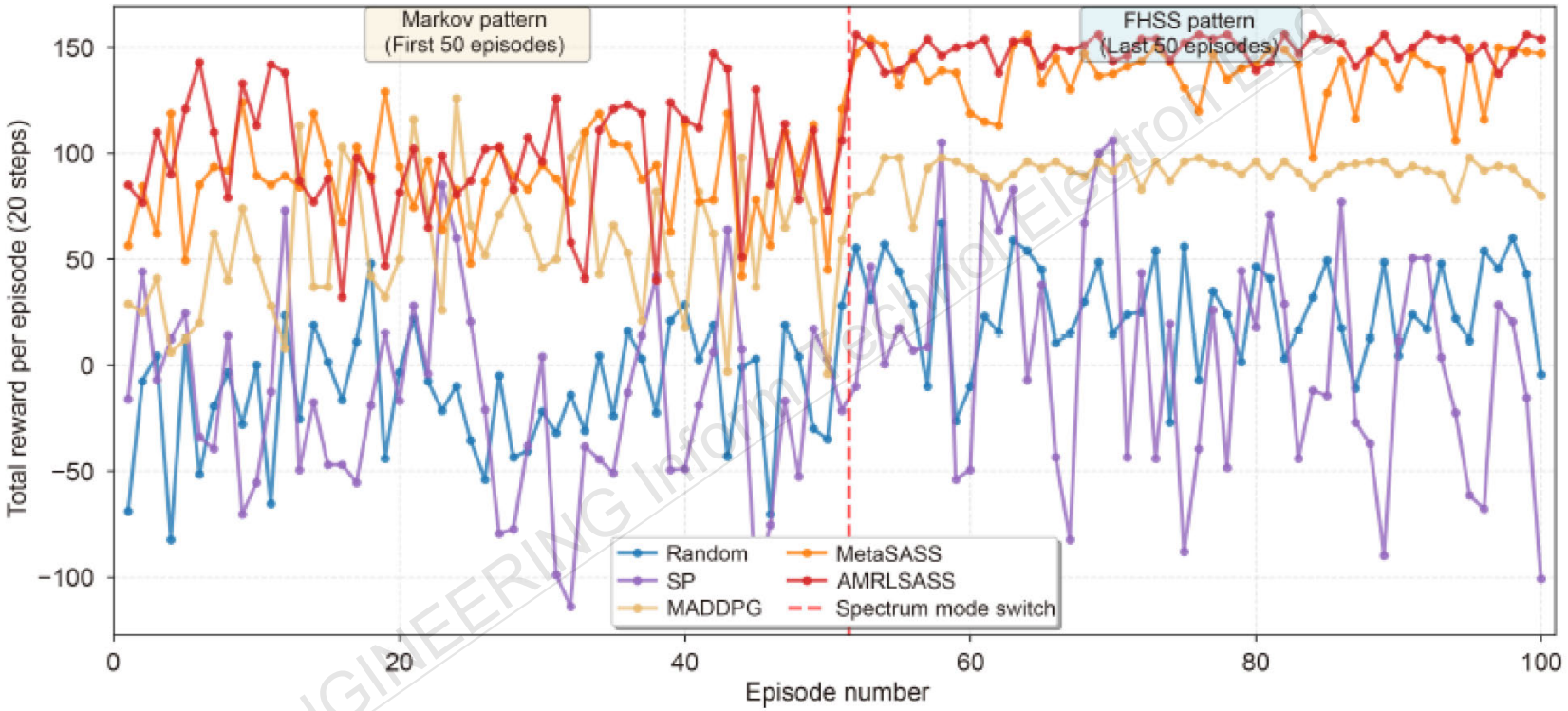


Fig. 5 Comparison of different algorithms in terms of generalization performance (episode-level)

ENGINEERING Information Technology Electron Ltd

Results

Table 1 Performance comparison of different algorithms

Algorithm	All training episodes		Final training episode	Last 50 training episodes	
	Average reward	Std. reward	Final reward	Average reward	Std. reward
Random	−8.62	25.99	−61	−5.63	24.93
SP	23.24	79.73	−129	9.69	81.46
MADDPG	88.21	40.31	99	113.69	21.88
MetaSASS	34.94	18.58	45.68	45.75	0.24
AMRLSASS	37.08	16.53	45.83	45.74	0.17

Table 4 Average reward comparison of different algorithms under the generalization test

Algorithm	Markov pattern		FHSS pattern	
	Episode reward	Step reward	Episode reward	Step reward
Random	−13.12	−0.66	26.90	1.34
SP	−18.09	−0.90	4.51	0.23
MADDPG	56.00	2.80	91.24	4.56
MetaSASS	88.67	4.43	137.99	6.90
AMRLSASS	97.86	4.89	149.50	7.48

Table 2 Performance comparison on the Markov pattern

Algorithm	Average reward	Urgent packet success rate (%)	Normal packet success rate (%)	Urgent packet timeout rate (%)	Normal packet timeout rate (%)	Channel utilization rate (%)	Average queue length	Channel conflict count
AMRLSASS	3.25	61.73	2.99	11.10	0	100	28.29	0
MetaSASS	3.04	60.51	3.66	12.77	0	100	29.21	0
MADDPG	0.29	47.93	2.40	21.20	20.88	80.19	30.62	20.88
SP	−2.08	39.31	0.92	30.42	27.00	70.72	28.54	27.00
Random	−2.85	35.38	1.69	32.11	33.62	61.54	30.00	33.62

Table 3 Performance comparison on the FHSS pattern

Algorithm	Average reward	Urgent packet success rate (%)	Normal packet success rate (%)	Urgent packet timeout rate (%)	Normal packet timeout rate (%)	Channel utilization rate (%)	Average queue length	Channel conflict count
AMRLSASS	4.55	58.89	1.41	5.55	0	100	33.83	0
MetaSASS	4.30	61.00	2.89	7.63	1.88	99.22	32.17	1.88
MADDPG	0.74	47.27	1.27	14.22	33.75	83.37	34.25	33.75
SP	−2.59	35.72	0.46	24.81	53.88	65.78	33.83	53.88
Random	−1.92	40.07	0.79	18.47	65.25	61.55	33.42	65.25

ENGINEERING Inform Technol Electron Eng

Conclusions

In this work, we propose a novel adaptive meta-reinforcement learning scheme based on a decentralized partially observable Markov decision process and dynamic task encoding, aiming to provide a rapid adaptation mechanism to achieve efficient spectrum access and service scheduling in complex maritime environments. The design of the adaptive meta-learning rate control can provide some inspiration for further research on intelligent decision-making in time-varying and unknown wireless networks.

ENGINEERING Inform Technol Electron Eng



Zhongyang MAO received the Ph.D. degree from Naval Aviation University, China, in 2011. He is currently a professor in Naval Aviation University, Yantai, China. His research areas include modern communication system and intelligent networking.



Jiahuan GENG received the B.E. degree from National University of Defense Technology, China, in 2011. He is currently pursuing the Ph.D. degree in Naval Aviation University, Yantai, China. His research area include intelligent networking and reinforcement learning.



Faping LU received the Ph.D. degree from Naval Aviation University, China, in 2021. He is currently an Associate Professor in Naval Aviation University, Yantai, China. His research areas include non-sinusoidal communication and intelligent networking.

ENGINEERING Inform Technol Electron Eng